PB86-213535

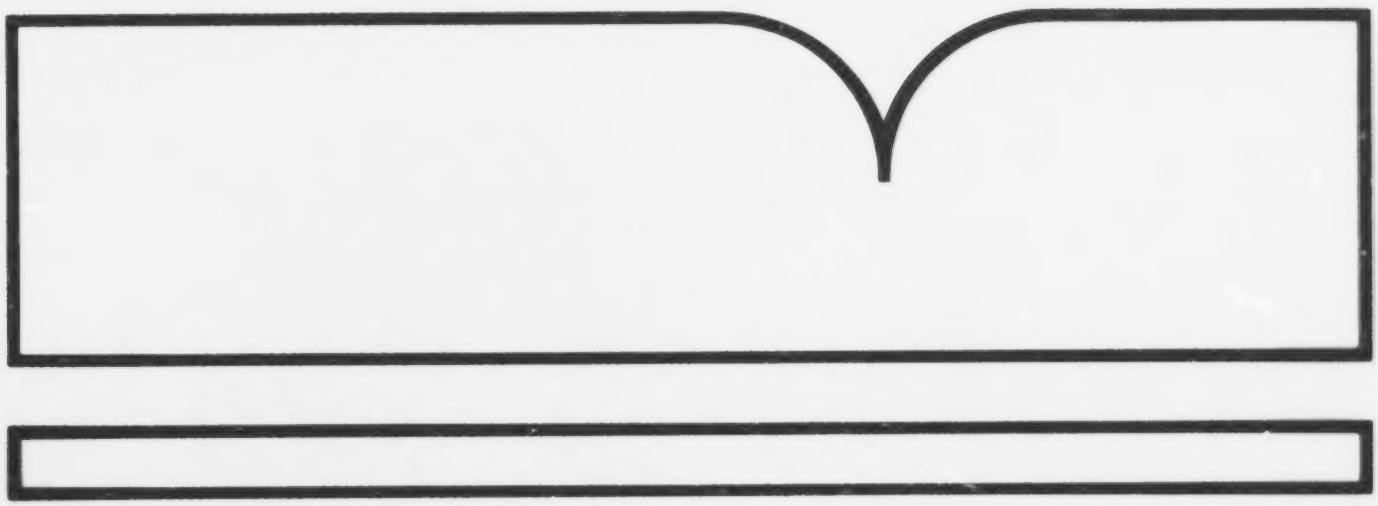
Limited Occupational Mortality Data and Statistics

North Carolina Univ. at Chapel Hill

Prepared for

Health Effects Research Lab.
Research Triangle Park, NC

1985



U.S. Department of Commerce
National Technical Information Service
NTIS

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA/600/D-86/111	2. PB86-213535
4. TITLE AND SUBTITLE LIMITED OCCUPATIONAL MORTALITY DATA AND STATISTICS	5. REPORT DATE 1985
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Michael J. Symons*, Donald L. Doerfler, and Yang C. Yuan	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Biostatistics and Occupational Helath Studies Group School of Public Health UNC, NC	10. PROGRAM ELEMENT NO.
	11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS Health Effects Research Laboratory-RTP,NC Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, NC 27711	13. TYPE OF REPORT AND PERIOD COVERED
	14. SPONSORING AGENCY CODE EPA/600/11

15. SUPPLEMENTARY NOTES
BIOSTATISTICS: Statistics in Biomedical Public Health and Environmental Sciences, edited by P. K. Sen, Elsevier Science Publ. B.V. (North-Holland), 1985

16. ABSTRACT

From principles of survival analysis statistical approaches are identified for a major problem with occupational mortality studies when few deaths, as with cause specific mortality, and no exposure data on the individuals are available. Specifically, this limitation is the general inadequacy of a geographic population as a comparison group for employed populations. Qualitative aspects of the so-called healthy worker effect are briefly presented, while the essential features of statistical strategies for part of this issue are sketched in detail. Although the occupation epidemiology literature focuses on selection bias as the source of the problems with the standardized mortality ratio, the potential inadequacy of the proportional hazards model for occupational mortality evaluations is emphasized here. A practical, nonproportional hazards model, patterned after the slope-ratio design in bioassay, may provide a more sensitive approach.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS (This Report)	21. NO. OF PAGES
RELEASE TO PUBLIC	UNCLASSIFIED	26
	20. SECURITY CLASS (This page) UNCLASSIFIED	22. PRICE

EPA Form 2220-1 (Rev. 4-77)

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

NTIS is authorized to reproduce and sell this report. Permission for further reproduction must be obtained from the copyright owner.

EPA/600/D-86/111

BIOSTATISTICS: Statistics in Biomedical, Public Health and Environmental Sciences, edited by P.K. Sen
© Elsevier Science Publishers B.V. (North-Holland), 1985

LIMITED OCCUPATIONAL MORTALITY DATA AND STATISTICS

Michael J. Symons, Donald L. Doerfler, and Yang C. Yuan

Department of Biostatistics and Occupational Health Studies Group
School of Public Health
University of North Carolina at Chapel Hill

From principles of survival analysis statistical approaches are identified for a major problem with occupational mortality studies when few deaths, as with cause specific mortality, and no exposure data on the individuals are available. Specifically, this limitation is the general inadequacy of a geographic population as a comparison group for employed populations. Qualitative aspects of the so-called healthy worker effect are briefly presented, while the essential features of statistical strategies for part of this issue are sketched in detail. Although the occupational epidemiology literature focuses on selection bias as the source of the problems with the standardized mortality ratio, the potential inadequacy of the proportional hazards model for occupational mortality evaluations is emphasized here. A practical, non-proportional hazards model, patterned after the slope-ratio design in bioassay, may provide a more sensitive approach.

KEY WORDS: Survival analysis, proportional hazards, standardized mortality ratio, non-proportional models, slope-ratio design, bioassay, occupational epidemiology

1. INTRODUCTION AND SUMMARY

In the evaluation of the cause specific mortality experience of an occupational cohort, statistical analysis options are limited with only a few observed deaths and when no exposure evaluations on individuals are available. The standardized mortality ratio (SMR) is the historical summary statistic for such studies. It is the maximum likelihood estimator for a presumed constant ratio of the employee's rates, at every age band, to those of the population of an appropriate geographic area which is composed of employable and unemployable individuals. The so-called healthy worker effect represents a major criticism of the SMR as a summary statistic, since employees are typically healthier than members of the general population. The unavailability of mortality rates for employees in the United States that may reasonably represent an unexposed group is a very fundamental limitation.

Two alternative single parameter models are suggested that circumvent some of the criticisms of the SMR, but require mortality rates of employable individuals at the ages of hire. The first is an SMR construction presuming the appropriate standard rates are those of a general population multiplied at each age band by the ratio of the rate for employables at the ages of hire to the corresponding

rate for a general population. This synthetic schedule of rates then is proportional to that of the general population at all ages as determined by mortality rates for an employed group at the ages of hire.

A second schedule of synthetic rates is suggested presuming a regression from the mortality rates at the ages of hire for employables to those at the ages of retirees in the general population. The rationale being that the mortality rates at the older ages are by definition those of the survivors of all younger ages, and is compatible with a dying off of those at greater risk in the general population during the younger ages. An SMR construction, presuming a proportional effect on mortality from exposure in the workplace, is a convenient choice. However, the non-proportional character of this synthetic standard may warrant a parameter reflecting this understructure. It appears that such a parameter may be more relevant to quantify an excess risk of the employed group than the proportional one estimated by the SMR based upon any standard schedule of rates. Further, it appears to accommodate the lack of fit frequently exhibited by the SMR, namely the tendency for SMRs to increase with age or cumulative exposure.

Another selection force operating on those employed is that of leaving the industry or moving from the jobs of hazardous exposures by the most sensitive individuals. Statistical aspects of competing risks modelling may be helpful in this regard, but this approach requires reliable data on the reason for an exit from an area or the reason for quitting the industry and better follow-up of short-term workers. Analysis experience with even limited data on these points may be the best guide to addressing statistically these complex issues.

Each of these qualitative criticisms of occupational mortality studies could be addressed statistically given the required data. It is in the quality and extent of data, rather than in development of statistical techniques, where the greater need is identified. For example, the adversary climate of government and industry seems to inhibit a good faith collection and retention of data by companies. Presently, there is no protection afforded to a company having data that incriminates itself even if prudent industrial hygiene practices were maintained in past times by its management. Further, similar conflicts of interest induce workers to understate personal exposures, for example, smoking level, thereby limiting the scientific apportionment of injury to occupational and non-occupational factors. Perhaps incentives could be devised to assure the collection and retention of quality data by industry and the cooperation of employees.

Attention is directed to a one parameter, non-proportional hazards model as an alternative to the constant rate ratio estimated by the SMR. Support in two areas

for this alternative summary statistic is developed. First, this model is patterned after a slope-ratio design from bioassay where age is a proxy for dose, which in occupational mortality applications can be viewed as highly correlated with cumulative exposure. Second, criticisms of the SMR can be seen as evidence for lack of fit of the proportional model, for example, in its tendency to increase when computed for older age bands, longer periods of hire, or greater accumulations of exposure. The problems exhibited by the SMR seemed to be described generally as selection bias in the occupational epidemiology literature. Although there is reason for this concern, it may be a lesser one than simply an inadequate proportional hazards model.

Brief discussion of the potential implications from the inadequacy of a proportional hazards model for the unknown carcinogenic process is also offered.

2. A PARTITION OF THE HEALTHY WORKER EFFECT

Selection biases pose difficulties for any analysis strategy of occupational mortality studies. As a total, McMichael, Haynes, and Tyroler (1975) termed them the so-called "healthy worker effect." Meaning, in order to be and to stay employed one must physically go to work. As the population of any geographic area contains some who cannot go to work, mortality rates for the United States, or any single state, are thought to be greater than for the employed population of the geographic area. However, quantification of this idea has not been clearly set out, probably due to the absence of an operational definition. As a step in this direction, consider the healthy worker effect as composed of two selective forces:

(i) Selection at Hire: Some individuals in the general population do not appear before, or get past, the industrial employment offices at the usual ages of hire. Self-imposed or real physical/medical limitations are obstacles to employment. This selection force should be considered stronger for some industries or companies.

(ii) Selection by the Workplace Environment: For those initially employed, the most sensitive individuals leave the industry due to physical or mental intolerance for the environment of the workplace. Exposures, as well as the stresses of holding a job, individually or in combination, may result in the exit of some individuals from the employment roles of some companies. A lesser effect, job changes to positions of less exposure or stress, may also be operative.

In the context of these two presumed selective forces, the required data are generally not available. Even so, there are statistical methods that are applicable. For the selection of those employed, cause specific mortality rates at the ages of hire for these workers form the basis of two strategies for construct-

ing synthetic schedules of more appropriate mortality rates. For the selection out of the industry of those most sensitive to a workplace environment, reliable data on the reason for exit and better follow-up of these individuals would permit the application of competing risks to assist in the evaluation of the mortality experience of an employed group. Only for selection at hire are the statistical alternatives considered in further detail.

3. PARSIMONIOUS MODELS OF MORTALITY FOR SELECTION AT HIRE

As frequently the interest of an occupational mortality study is the risk from a specific cause of death, there may be few deaths upon which to estimate the hazard. Recall that with the constant hazard, θ, of an exponential model, the observed information available in the sample is proportional to the observed number of failures. Specifically,

$$\ell nL(\theta) = \sum_{i=1}^{n} \{\delta_i \ell n\theta - \theta t_i\} \qquad (1)$$

is the log-likelihood for n exponential observations t_i, where δ_i indicates a failure (δ_i=1) or a censored observation (δ_i=0) at t_i. The negative of the second partial of the log-likelihood is the observed information; visually,

$$-\frac{\partial^2 \ell nL(\theta)}{(\partial\theta)^2} = \sum_{i=1}^{n} \delta_i/\theta^2 . \qquad (2)$$

Consequently when the number of failures is small, only the simplest models should be considered since the observed information is proportional to the number of failures in the sample. Applications requiring an evaluation of a specific occupational mortality cause of death will be of special interest.

A. Standardized Mortality Ratio

Let t_i be the follow-up time for the i^{th} member of an occupational cohort, whose hazard can be written as

$$\lambda_i(t) = \theta R_i(t), \qquad (3)$$

where $R_i(t)$ is a known cause specific death rate appropriate for the i^{th} worker at time t. For $0 \le u \le t_i$, $R_i(u)$ is the pattern of standard rates relevant for the pertinent demographic factors, e.g., sex, race and calendar year of each achieved age. The above multiplicative form is termed a proportional hazards model, following Cox (1972). While at risk to a specific cause of death, the survival function for a worker surviving to time t_i, i.e., being censored by the end of follow-up, being lost during study, or by dying of another cause, is given by

$$S[t_i; \theta, R_i(u)] = \exp\{-\int_0^{t_i} \theta R_i(u)\, du\}. \qquad (4)$$

The log-likelihood for θ then is given by

$$\ell nL(\theta) = \sum_{i=1}^{n} \{\delta_i \ell n[\theta R_i(t_i)] - \theta \int_0^{t_i} R_i(u)du\}. \tag{5}$$

By direct calculation,

$$\frac{\partial \ell nL(\theta)}{\partial \theta} = \sum_{i=1}^{n} \{\delta_i/\theta - \int_0^{t_i} R_i(u)du\} \tag{6}$$

and then

$$\hat{\theta} = \sum_{i=1}^{n} \delta_i / \sum_{i=1}^{n} \int_0^{t_i} R_i(u)du = O/E, \tag{7}$$

the ratio of the total observed number of deaths from the specific cause to the total "expected" deaths under the schedules of rates $R_i(\cdot)$, for $i=1,2,\ldots,n$. These are expected, in that as Breslow (1978) has shown,

$$E(\delta_i) = E\{\int_0^{t_i} R_i(u)du\}, \tag{8}$$

where the length of follow-up is infinite. The estimator (7), SMR = O/E, was presented by Kilpatrick (1962) as the maximum likelihood estimator of the unkonwn constant ratio of the study to standard population specific rates. He used a Poisson model and a person-years summary of these same data. Breslow (1975, 1976, 1978) exhibited the SMR as the maximum likelihood estimator for the above survival analysis presentation of these data.

However, in actual cohort mortality studies the equality (8) holds only approximately, since in finite length studies follow-up losses are certain. Consequently, for common causes of death with even a short follow-up period or for more rare causes but with a long follow-up period, the inequality with equation (8) can be substantial. The approximation of the survival function (4) by the cumulative of the hazard (3) is essence of the difficulty. The Appendix contains elaboration on this point, as well as illustration of the differences in two alternative calculations of expected numbers, represented by each side of equation (8). Each has been used in practice. The result is to stress the use (7) for estimation of the common age specific rate ratio, although the sum of the expected values of the δ_i perhaps is more properly called the expected number of deaths. Further, this is a good example where following likelihood principles results in clarification of statistical practice.

For estimation of θ and testing of the null hypothesis that θ is one, the observed information in the data is

$$-\frac{\partial^2 \ell nL(\theta)}{(\partial\theta)^2} = \sum_{i=1}^{n} \delta_i/\theta^2. \tag{9}$$

Noting (9) and (8), the expected information reflects the observed and expected number of deaths, respectively. The expected information in the sample increases directly with the number of individuals n and with the follow-up time t_i for each individual in the cohort. The integral in (8) can be approximated as follows:

$$\int_0^{t_i} R_i(u)du \doteq \sum_{j=1}^{J_i} R_i(u_j) + f_i(J_i+1)R_i(u_{J_i+1}), \qquad (10)$$

where $R_i(u_j)$ is the constant, cause specific death rate in the standard population for follow-up year u_j and restricted to the relevant demographic characteristics of individual i, who is followed for J_i complete years and for fraction $f_i(J_i+1)$ of the last year. Fractions $f_i(j)$ for each year could be introduced if only portions of years, or none of some, were considered at risk. A common practice is to let $R_i(\cdot)$ be constant for five year intervals rather than single years and take $f_i(J_i+1)$ equal to one-half. These details and variations are reviewed by Symons and Taulbee (1984).

Breslow (1975) showed that Rao's efficient score test for H_o: $\ell n\theta=0$ is given by $(O-E)^2/E$. Likelihood ratio or Wald's criteria might also be used, but they are not so simple computationally. For hand calculations, the score statistic is convenient and can be expected to have power comparable to that of likelihood ratio and Wald test statistics, as shown by Lee, Harrell, Tolley and Rosati (1983) for proportional hazard models with covariables.

B. Two Synthetic Standard Constructions

Within the context of a proportional effect on the occupation mortality cause with biological relevance, an SMR type estimator of this effect is one approach for using any synthetic standard. Each of the two standards presented attempts to address the concern for a reduced risk of death at a typical age of hire by using a presumed available estimate of the age-at-hire death rate among employees for the cause of interest. One standard is proportional to the general population at all ages in the same proportion as the mortality rate at the age of hire for those employed to that for those in the general population. The second is a regression of the mortality rates at the age of hire for those employed to that of the general population at an age of retirement. As the latter synthetic standard is a non-proportional pattern, another summary statistic is also presented that may be more sensitive than the SMR, especially if the effect of the workplace environment is progressively greater at older ages, as would be consistent with a mortality rate that increased with cumulative exposure. The age at which death occurs is incorporated into this estimator, circumventing one criticism of the SMR; see Gaffey (1976). Use of the age pattern of deaths may be appropriate for some applications, for example with long, low level exposures, which may provide a more

powerful test statistic for this bioassay inspired alternative to the SMR.

1. Synthetic Standard Proportional to Rates of Hirees

Let $R_i^*(x_o)$ be the cause specific rate for employees at the ages of hire in the study cohort, or other comparable group of employed individuals, perhaps obtained from Social Security files as suggested by Goldsmith (1975). These may be available from the occupational mortality literature, but unfortunately such reporting is not customary; Pell, O'Berg and Karrh (1978) is an illustrative exception. The age band 15-25 years of age may be required to provide a reliable estimate for an individual with comparable demography, for example, having the same sex and race, and using $x_o = 20$ as a representative age of hire.

The calculation of an SMR estimate (7), using an approximation following (10), would be accomplished by replacing the general population schedule of rates $R_i(x_j)$ with

$$R_i^P(x_j) = [R_i^*(x_o)/R_i(x_o)]R_i(x_j) \ , \qquad (11)$$

where x_j is a representative age for the j^{th} age band. In Figure 1, the schedule of $\ln[R_i^P(x_j)]$ is presented approximately by the longer dashed lines at the logarithm at the midpoints of the 10 year age bands from age 15 through age 75. Although the pattern of death rates in this age range is largely exponential in nature, observed rates at the upper age bands are usually slightly less than those predicted by an exponential trend. The fit was improved by plotting against log-age, i.e., fitting a Weibull rather than Gompertz hazard as preferred by Cook, Doll and Fellingham (1969). Also, fitting an additive constant to a general exponential trend with age, as described by Makeham (1860), improves the representation of general population rates, but considerably complicates the determination of the synthetic standard.

With the availability of a mortality rate for the employed segment of the population at an age of hire, a more suitable standard schedule of rates may be prepared. The first synthetic schedule is motivated by the very essence of the SMR itself, namely, that the effect of the workplace on the mortality rates of its employees is in constant proportion at all ages. Visually this is portrayed in Figure 1 as a constant difference between the Smoothed U.S. Standard and the Proportional Synthetic Standard when the corresponding logarithms of the rates are plotted against log-age.

2. Synthetic Standard Regressing to U.S. Rates of Retirees

At the ages of hire a general population may have mortality rates in excess of those in a demographically similar employed group. (However, the previous

synthetic standard, nor the one to be described, requires this difference necessarily to be positive.) The thinking is that the unemployable portion in the general population is subject to greater death rates. With aging the tendency would be for these individuals to be depleted from the population more quickly than those initially employed and working in reasonably safe conditions. Consequently, the death rates of those survivors at the ages of retirement in the general population would then be more like those initially employed and surviving to the same ages of retirement than at the ages of hire. If every workplace environment posed serious threats to the longevity of its employees, this line of argument fails. In such harsh workplace environments, the survivors to retirement age would represent a very hardy group of individuals.

The second synthetic standard is a schedule of rates that progresses smoothly from reduced rates at the ages of hire for employed individuals to rates at a specified age of retirement for the general population survivors. Such a regression strategy is not new. Yates (1934) suggested a similar regression-to-the-mean phenomena in examining the pulmonary function of workers employed in dusty environments as a means to address the presence of workers at hire with the superior lung function and tolerance to irritant dust.

One simple way of trending the rates from those of the employed at the ages of hire to those of the general population survivors is to use a linear pattern in the death rates over log-age as depicted by the shorter dashed line in Figure 1 for the Non-Proportional Synthetic Standard. This construction implies greater rates than the U.S. Standard at ages past retirement, but seems useful at least as a first approximation. Therefore, a slightly less steep slope is probably more appropriate for the unexposed pattern of rates, but then the slope described should be somewhat conservative. Also, due to the decrease in reliability of death certificate coding of cause of death at the older ages, an SMR calculation would be terminated at this age band of retirement where the regression ends.

Under proportional hazards the observed over expected ratio (7) is a desirable estimator, utilizing the exponential of the logarithm of the rates in Figure 1 for the Non-proportional Synthetic Standard to obtain the expected number of deaths. However, by this latter construction a parameter reflecting a non-proportional effect of a detrimental workplace environment may be more relevant than the proportional one given in (3). For example, if the effect increases with cumulative time in the workplace environment, a descriptive one parameter model is a slope. An estimate is obtained by fitting a line to the logarithm of the observed age specific death rates, but required to go through the logarithm of the death rate at the log-age of hire in the employed group. There is an analogy to the bioassay

design with the power dose metameter in a slope-ratio design; see Finney (1964). In this application, advancing age is a proxy for cumulative exposure. The null hypothesis is that the slope in the logarithm of death rates across log-age is given in Figure 1 by

$$\beta_0 = \{\ln[R_1(70)] - \ln[R_1^*(20)]\}/[\ln(70) - \ln(20)]. \quad (12)$$

In the presence of detrimental effects to mortality from the workplace environment, alternative values of the slope parameter are greater than β_0.

Following Frome (1983), maximum likelihood estimation of the slope descriptive of the experience of the cohort could be obtained from the Poisson regression likelihood

$$L(\beta) = \prod_{j=0}^{J} \{e^{-\lambda_j(\beta)P_j}[\lambda_j(\beta)P_j]^{d_j}\}, \quad (13)$$

where d_j is the observed number of deaths in the j^{th} age band; P_j is the number of person-years of experience accrued in age band j; and

$$\lambda_j(\beta) = R^*(e^{a_0}) \exp\{\beta(a_j - a_0)\}, \quad (14)$$

with a_j being the logarithm of the midpoint age of the j^{th} age band, i.e., $x_j = \exp(a_j)$. The subscript i has been suppressed, indicating that only one race-sex group is being considered. Differentiating the logarithm of $L(\beta)$ with respect to β yields the following equation that would require iterative methods to obtain the maximum likelihood estimate, $\hat{\beta}$, satisfying:

$$R^*(e^{a_0}) \sum_{j=1}^{J} (a_j - a_0)\, e^{\hat{\beta}(a_j - a_0)} P_j = \sum_{j=1}^{J} (a_j - a_0) d_j. \quad (15)$$

The null hypothesis value β_0 in (12) is a natural initial estimate and convergence to $\hat{\beta}$ should not be difficult by a Newton-Raphson algorithm.

A significance test of the null hypothesis $H_0: \beta = \beta_0$ versus the alternative hypothesis $H_a: \beta > \beta_0$ can be approximated by

$$z_s = (\hat{\beta} - \beta_0)/\sqrt{\hat{V}(\hat{\beta})}, \quad (16)$$

where the variance of the maximum likelihood estimator for β can be approximated by the inverse of the negative of the second partial of the log-likelihood evaluated at $\hat{\beta}$, specifically

$$\hat{V}(\hat{\beta}) = \{R^*(e^{a_0}) \sum_{j=1}^{J} (a_j - a_0)^2 e^{\hat{\beta}(a_j - a_0)} P_j\}^{-1}. \quad (17)$$

A value of the test statistic exceeding the $(1-\alpha)100$ percentile of the standard normal distribution supports rejecting $\theta=\theta_0$, which would be appropriate when the workplace conditions are detrimental. A score statistic will not have its usual simplifications since θ_0 is not zero, but it or the likelihood ratio statistic could be computed in place of the Wald statistic (16).

4. EXAMPLE

The slope-ratio model and variations of a proportional hazards model are illustrated with data originally in Dement (1980) and updated in Dement, et al. (1983). For white males working one or more months between 1940 and 1975 with chrysotile asbestos, the numbers of deaths, d_j, for lung cancers (ICDA 162-163) and the person-years, P_j, of work experience accrued were summarized by 10 year age bands: 15-24, 25-34, ..., 65-74, and 75-84. Indexing these age bands by j, the a_j's are the natural logarithm of the midpoints of these age bands, namely, $a_0 = \ln(20)$, $a_1 = \ln(30)$,..., and $a_6 = \ln(80)$. [With such a long study period, the numbers of deaths and person years could be tabulated by age bands for shorter calendar year periods. General population age specific lung cancer death rates would then be employed as described by Monson (1974)]. In addition, the lung cancer death rate for an employed group at the age band of hire, denoted by $R^*(20)$, was approximated by the data in Pell, O'Berg and Karrh (1978).

These data are exhibited in Table 1 and the various standards are graphically displayed in Figure 1. The logarithm of the observed age specific lung cancer rates are also plotted as open squares in Figure 1. The proportional models can be projected onto Figure 1 as a line parallel to one of the standards or as points at a constant distance above (for an excess) the points determining the straight line smoothings in Figure 1. The amount of separation from the standard is determined by the observed rates. The slope-ratio model parameter is the slope of a line with intercept at $[\ln(20), \ln R^*(20)]$ and whose magnitude is determined by the observed rates.

Test statistics were computed for each of three models using data in Table 1. Two variations on proportional models were included. The slope-ratio model is determined by the age specific hazard (14). Each model has one parameter in the generic Poisson log-likelihood

$$\ln L(\theta) = \sum_{j=0}^{J} \{-\lambda_j(\theta)P_j + d_j \ln[\lambda_j(\theta)P_j]\} . \qquad (18)$$

The three models can be specified by their use of one parameter in the age specific hazard as follows:

$$\lambda_j(\theta=\gamma) = \exp(\gamma)\, R(e^{a_j}) \qquad (19)$$

Table 1. Lung Cancer Deaths and Person-Years of Experience by Ten Year Age Bands for White Male Chrysotile Asbestos Textile Workers, 1940-1975, and 1960 U.S. General Population for White Males

Age Band	Person Years	Lung Cancer Deaths ICDA 162-163	Observed Rate	1960 U.S. Male Standard†		
				Population	Deaths ICDA 162-163	Rate
15-24	4023.1	0	(0.000 000 928)*	10,500,224	13	0.000 001 238
25-34	9039.2	0	0.000 000	9,904,838	132	0.000 013 326
35-44	9645.6	0	0.000 000	10,524,984	1029	0.000 097 767
45-54	7009.8	16	0.002 283	9,147,704	4480	0.000 489 740
55-64	2743.9	11	0.004 009	6,883,578	9536	0.001 385 326
65-74	606.5	8	0.013 189	4,606,007	9760	0.002 118 972
75-84	71.7	0	0.000 000	1,804,478	3195	0.001 770 595

*Approximated as 75% of the corresponding U.S. lung cancer rate, following the age specific mortality ratios in Table 3 and lung, bronchus and trachea rates in Figure 3 of Pell, O'Berg and Karrh (1978).

†Population counts are from 1960 Census of Population, Characteristics of the Population. Deaths are from 1960 Vital Statistics of the United States, Volume II: Mortality.

Source: Dr. John M. Dement kindly provided the person-years and numbers of lung cancer deaths for this purpose.

is the first model (Model I), the familiar SMR model described earlier with equation (3). The second model (Model II) is

$$\lambda_j(\theta=\alpha) = \exp(\alpha)\, R^*(e^{a_0}) \exp[B(a_j-a_0)], \tag{20}$$

where the slope B was determined by a least squares fit of a line to the points $[a_j, \ell nR(e^{a_j})]$. A graphical fit should be sufficient, however, as can be seen in Figure 1. It uses the smoothed U.S. Standard rates and is required to pass through the point $[a_0, \ell n R^*(e^{a_0})]$. This is the Proportional Synthetic Standard in Figure 1 and follows from a parallel line bioassay design. The third model (Model III) is given by

$$\lambda_j(\theta=\beta) = R^*(e^{a_0}) \exp[\beta(a_j-a_0)], \tag{21}$$

the slope-ratio bioassay alternative. Two other proportional models not included are variants of (19). The first replaces $R(\cdot)$ with the proportional standard given in (11) and the second replaces $R(\cdot)$ by (21) with β specified as

$$\beta_0 = \{\ell nR(e^{a_5}) - \ell nR^*(e^{a_0})\}/(a_5-a_0). \tag{22}$$

Notice for this application, the data in 75-84 year age band were not used.

The Rao-Score, Wald, and likelihood ratio strategies each determines approximately a one degree of freedom chi-square test statistic. Each is described generically as follows using θ to represent the single parameter of each model as in equation (18) for $\ell nL(\theta)$. Let

$$U(\theta) = \frac{\partial \ell nL(\theta)}{\partial \theta} \tag{23}$$

denote the first partial of the log-likelihood with respect to the parameter of the model and

$$I(\theta) = -\frac{\partial U(\theta)}{\partial \theta} = -\frac{\partial^2 \ell nL(\theta)}{(\partial\theta)^2} \tag{24}$$

Now the three test statistics can be specified as follows:

(i) The Rao-Score (RS) one degree of freedom chi-square is

$$\chi^2_{RS} = [U(\theta_0)]^2/I(\theta_0)\ , \tag{25}$$

where θ_0 is the null hypothesis value of the model parameter. For the first and second models, γ_0 and α_0 are zero, respectively, and for the slope-ratio model the null hypothesis value of β is approximated by β_0 in (22).

(ii) The Wald Statistics (WS) one degree of freedom chi-square is

$$\chi^2_{WS} = (\hat{\theta}-\theta_0)^2/\hat{V}(\hat{\theta})\ , \tag{26}$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ. Specifically, $\hat{\theta}$ is the value of the parameter θ at which $\ell nL(\theta)$ is maximized. This requires that $U(\hat{\theta}) = 0$ and $I(\hat{\theta})$ be non-negative. Also, $V(\hat{\theta}) = [I(\hat{\theta})]^{-1}$.

(iii) The Likelihood Ratio (LR) statistic is also a one degree of freedom chi-square and

$$\chi^2_{LR} = -2[\ell nL(\theta_o) - \ell nL(\hat{\theta})]. \qquad (27)$$

The test statistics, parameter estimates, and their estimated standard errors are given in Table 2 for the chrysotile asbestos cohort data in Table 1. Clearly the test statistics for each model reject the null hypothesis value for its corresponding model parameter. There is very strong evidence for an excess of lung cancer mortality among these workers. The nature of the patterns of this excess is the important issue. These results are suggestive of two points worthy of further study. First, directly incorporating information about the perceived lower rate of lung cancer among employees than among the general population at the

Table 2. Analysis Results from Three Models for the Chrysotile Asbestos Data

		Parameter			Test Statistics		
Model	Description	θ_o	$\hat{\theta}$	Standard Error	Rao-Score	Wald Statistic	Likelihood Ratio
I	Constant rate ratio	$\gamma_o=0$	1.295	0.1690	67.4	58.7	39.8
II	Intercept	$\alpha_o=0$	1.706	0.1690	129.1	101.9	62.1
III	Slope	$\beta_o=6.17$	7.816	0.1575	140.6	108.8	64.5

ages of hire can increase the statistical leverage of the proportional hazards model. Is the pattern of excess being measured from too high a standard? The test statistics for Model II are about double those for Model I. Second, the slope-ratio may be more sensitive than a proprotional hazards approach. The test statistics for Model III are also about double those for Model I. Is the excess lung cancer proportional at all ages, or progressively greater at older ages? The issue identified is a choice between models. A resolution is difficult without biologic understanding of the cancer mechanism; other alternatives to proportional hazards should also be sought.

5. A BIOASSAY PERSPECTIVE

The essential difficulty with occupational mortality studies can be viewed from a bioassay perspective. The discussions of the synthetic standard options in the previous subsections attempt to identify a pattern over age for the unexposed but employed individuals. The synthetic standards correspond to the responses for the standard preparation in bioassay, with the dose metameter corresponding to age in the analysis of occupational mortality studies. However, there are usually no data for the employed but unexposed group. In the absence of any data, analysis chcices are largely subjective and consequently may be difficulty to defend.

The two proposed synthetic standards depend primarily on the addition of one data point, namely, the cause specific mortality rate for employed individuals at the age of hire, hopefully providing a more appropriate schedule of rates for the unexposed employees than that by using the general population specific rates. That such a schedule of rates is appropriate for unexposed employees makes sense only with an assumption that the effects of exposure on mortality are not immediate. This is probably reasonable for many causes of death, especially those resulting from chronic exposure, but the assumption is not always reasonable, for example with accidental deaths.

To summarize, Table 3 contains a correspondence between features of an occupational mortality evaluation with those of a bioassay, where statistical models and procedures are long-standing. The main observation is that data on the comparison group are generally not available from occupational mortality studies. The synthetic standards were devised to partially remedy this shortcoming. With each synthetic standard, a proportional model and the corresponding observed over expected ratio estimator (3) is one analysis strategy. However, the analogue to the slope-ratio model from bioassay, suggests a reasonable alternative with biologic support, namely, a detrimental effect that increases with cumulative exposure and is measured in proxy by age.

6. DISCUSSION AND CONCLUSION

An association between the way of constructing the synthetic standards and the summary statistic for the underlying bioassay model raises at least two questions. First, is a proportional hazards model appropriate? If a non-proportional hazard seems to fit occupational mortality data better than a proportional one, are there any implications about the basic carcinogenic process? Second, are there other possibilities for the rates of the unexposed employees?

Table 3. Comparison of Bioassays and Occupational Mortality Evaluations

Aspect	Type of Study	
	Bioassay	Occupational Mortality
1. Nature of Study	Experimental	Observational
2. Groups	Test/Standard Preparations	Exposed/Unexposed Employee Groups
3. Data on	Both Groups	Exposed Only*
4. Response Trend on	Dose	Time Employed or Age
5. Parallel Model	Parallel Line Assay	Proportional Rates
a. Group Differences	Estimated	Relative to Known Pattern
b. Common Slope	Estimated	Assumed Known
c. Linearity	Testable	Not Strictly Needed
6. Non-Proportional Model	Slope-Ratio	Alternative Slope
7. a. Group Difference	Estimated	Relative to Known Slope
b. Common Intercept	Estimated	Assumed Known
c. Linearity	Testable	Assumed

*That there are no data on the employed, strictly unexposed, group is the main limitation of occupational mortality studies.

Regarding the relevance of the proportional hazards model, most of the references to the SMR statistic are very critical; see especially Gaffey (1976) and Wong (1970). Symons and Taulbee (1981) review several of the criticisms of the SMR and provide suggestions for improving its appropriateness. Unfortunately, their suggestion to restrict the age range to improve its applicability requires medical input on the natural history of the disease. Breslow (1975, 1976, 1977) presents the SMR as the maximum likelihood estimator for a proportional hazards model, but he does not consider goodness of fit. However, he mentions using direct rates rather than the indirect adjustment of the SMR approach. The proportional and non-proportional approaches can each be viewed as fitting the observed rates and then testing for a common intercept or a common slope, respectively. Breslow (1983) provides a recent comprehensive review, including Poisson regression of which likelihood (18) is an example. However, his paper is oriented towards proportional hazards and applications where data on the accumulated exposure are available. Again, the common situation where data are very limited is of primary concern here.

Empirically the support for the proportional hazards model is scattered. Amandus (1982) reports on the analysis of the mortality experience of a cohort of coal

mine workers and notes the non-parallelism in some cumulative hazard plots. However, he does not include a quantitative assessment. Dement (1980, 1982) assembled and reported the results of the analysis of the lung cancer mortality of a cohort of chrysotile workers. Dement (1983) contains the updated data used for the example. Gaynor (1983) provides further modelling and analysis of the same data. He presents non-parametric cumulative hazard curves that suggest deviation from the proportional hazards model, but with so few deaths beyond age 65, a clear violation is not reported. Visually, the cumulative hazard plots in Figure 4.1 of Gaynor (1983) suggest that the pattern of increased risk for all exposed groups would not be inconsistent with the non-proportional hazard in (14). For an application of evaluating fertility in an occupational setting, Boyle (1983) used an external standard for birth rates, as in (3), and found that this proportional model was inadequate.

In addition, there is considerable evidence for lack of fit to occupational mortality by the proportional hazards model in the critical comments surrounding the SMR. Although usually attributed to unestimable biases as part of the healthy worker effect, another interpretation is that the ratio of the study to standard age specific rates simply tends not to be constant. For example, McMichael (1976) notes the tendency of the SMR to increase when restricted to older ages. Gilbert (1982) reviews the occupational epidemiology literature in this regard and notes that the SMR tends to increase with longer exposure, longer employment and older age. The non-proportional model (14) could accommodate these phenomena observed with the SMR. When exposure data are not available and using age as a proxy for cumulative exposure, model (14) can be viewed as a slope-ratio design from bioassay when the pattern in the unexposed group is considered known.

Brief comments are made concerning the implications of non-proportionality for the underlying cancer risk. Breslow (1977) remarks that the multiplicative effect on the hazard due to membership in the occupational cohort is implied by multistage theories of carcinogenesis in which the last stage is affected by the occupation exposure. Whittemore (1977) establishes this relationship. If a non-proportional model seems more appropriate for occupational mortality data of site specific cancers, then either the multistage theory is inappropriate or other stages than the last are affected by the occupational exposure. The fit of proportional hazards model has implications to the underlying carcinogenic theory and, perhaps, for undue concern about selection bias in the occupational epidemiological literature. Critical evaluation of its fit is needed.

Regarding other schedules of standard mortality rates suggested by Figure 1, there are two in addition to the pair of synthetic schedules described above.

They are the smoothed U.S. general population death rates and another synthetic one with slope β_0 in (12) but originating at the logarithm of the general population death rate at log-age a_0. The first is usual choice for standard rates and was used by (19). The latter is more of a logical possibility, however.

Finally, it is helpful to note that if data on exposure for each employee are available, even an indication of categorically "exposed" or "not exposed," internal comparisons are then possible. The Cochran-Mantel-Haenszel procedure was set out by Cochran (1954) and was refined by Mantel and Haenszel (1959) to accommodate very slim two-by-two tables. It is applicable to many problems, particularly relevant is the survival application described by Mantel (1966). Gilbert (1979), 1982) and Gilbert and Marks (1979) demonstrated its applicatility to occupational mortality with their analysis of the mortality experience of workers exposed to radiation. Breslow (1975, 1983) indicated and also endorsed the use of internal standards, where the pooled experience of the exposure groups becomes the standard schedule of rates under the null hypothesis. Also, Woolson (1981) showed that the indirect adjustment procedure, of which the SMR is an example, can be viewed as a limiting case of the Cochran-Mantel-Haenszel procedure when the unexposed group becomes very large relative to the exposed group.

The point of mentioning the Cochran-Mantel-Haenszel procedure is to illustrate that statistically powerful methods are available for the evaluation of the mortality experience of an occupational cohort with appropriate data. With even more data on the level of exposure than just whether an individual was exposed or not, the time dependent covariates sketched by Breslow (1983) and utilized by Gaynor (1983) can be incorporated into the principles of survival analysis to utilize exposure by calendar year worked, for example. However, critical evaluation of the proportional hazards assumption in those cases still seems prudent.

ACKNOWLEDGEMENTS: This work was started in 1980 while MJS was on leave with the Health Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. Various subsequent fragments of support have come from the Epi-Stat Unit of the Cancer Research Center, and from the Occupational Health Studies Group, School of Public Health, University of North Carolina at Chapel Hill. This research support is gratefully acknowledged.

REFERENCES:

[1] Amandus, H., A comparison of analytical techniques for occupational mortality studies with an empirical example. Ph.D. Dissertation in Epidemiology. University of North Carolina at Chapel Hill (1982).

[2] Bittinger, C.S., A comparison of two methods of calculating the expected number of deaths for the standardized mortality ratio. Report for Master of Science in Public Health Degree in Biostatistics. University of North Carolina at Chapel Hill (1982).

[3] Boyle, K.E., Survival model for fertility evaluation. Ph.D. Dissertation in Biostatistics. University of North Carolina at Chapel Hill (1983).

[4] Breslow, N.E., Analysis of survival data under the proportional hazards model, International Statist. Rev. 43:45-58 (1975).

[5] Breslow, N.E., Some statistical models useful in the study of occupational mortality. Environmental Health: Quantitative Methods. Proceedings of a Conference on Environmental Health, Whittemore, A. (ed.). Philadelphia, SIAM: 88-103 (1977).

[6] Breslow, N.E., The proportional hazards model: applications in epidemiology, Comm. in Statist.:Theory and Methods A7(4):315-332 (1978).

[7] Breslow, N.E., Lubin, J.H., Marek, P., and Laugholz, B., Multiplicative models and cohort analysis, J. Amer. Statist. Assoc. 78:1-12 (1983).

[8] Chiang, C.L., On constructing current life tables, J. Amer. Statist. Assoc. 67:538-541 (1972).

[9] Cochran, W.R., Some methods for strengthening the common chi-square tests Biometrics 10:417-450 (1954).

[10] Cook, P.J., Doll, R., and Fellingham, S.A., A mathematical model for the age distribution of cancer in man, International J. of Cancer 4:93-112 (1969).

[11] Cox, D.R., Regression models and life tables (with discussion), J. Roy. Statist. Soc. Series B, 34:187-220 (1972).

[12] Dement, J.M., Estimation of dose and evaluation of dose-response in a retrospective cohort mortality study of chrysotile asbestos textile workers. Ph.D. Dissertation in Environmental Sciences and Engineering. University of North Carolina at Chapel Hill (1980).

[13] Dement, J.M., Harris, Jr. R.L., Symons, M.J., and Shy, C.M., Estimates of dose-response for respiratory cancer among chrysotile asbestos textile workers, Ann. of Occup. Hygiene 26:869-887 (1982).

[14] Dement, J.M., Harris, R.L., Symons, M.L., and Shy, C.M., Exposures and mortality among chrysotile asbestos workers. Part II: Mortality, Amer. J. of Industrial Med. 4:421-433 (1983).

[15] Enterline, P.E., Not uniformly true for each cause of death, J. of Occupational Med. 17:127-128 (1975).

[16] Frome, E.L., The analysis of rates using Poisson regression models, Biometrics 39:665-674 (1983).

[17] Gaffey, W.R., A critique of the standardized mortality ratio, J. of Occupational Med. 18:157-160 (1976).

[18] Gaynor, J.J., A framework that incorporates repeated measurements into the hazard. Ph.D. Dissertation in Biostatistics. University of North Carolina at Chapel Hill (1983).

[19] Gilbert, E.S., The assessment of risks from occupational exposures to ionizing radiation. Energy and Health, Breslow, N.E. and Whittemore, A. (eds.), Society for Industrial and Applied Mathematics, Philadelphia: 209-225 (1979).

[20] Gilbert, E.S., Some confounding factors in the study of mortality and occupational exposures, Amer. J. Epidemiol. 116:177-188 (1982).

[21] Gilbert, E.S. and Marks, S., An analysis of the mortality of workers in a nuclear facility, Radiation Research 79:122-148 (1979).

[22] Goldsmith, J., What we expect from an occupational cohort? J. Occupational Med. 17:126-217 (1975).

[23] Goldsmith, J.R. and Hirshberg, M.S., Mortality and industrial employment, J. Occupational Med. 18:161-164 (1976).

[24] Hill, I.D., Computing man years at risk, Brit. J. Preventive and Social Med. 26:132-134 (1972).

[25] Kilpatrick, S.J., Occupational mortality indices, Population Studies 16: 175-182 (1962).

[26] Lee, K.L., Harrell, F.E., Jr., Tolley, H.D., and Rosati, R.A., A comparison of test statistics for assessing the effect of concomitant variables in survival analysis, Biometrics 39:341-350 (1983).

[27] Levine, R.M., Symons, M.J., Balogh, S.A., Arndt, D.M., Kawsandik, N.T. and Gentile, J.W., A method to monitor the fertility of workers. 1. Method and pilot studies, J. of Occupational Med. 22:781-791 (1980).

[28] Makeham, W.M., On the law of mortality and the construction of annuity tables, J. Instit. of Actuaries 8 (1960).

[29] Mantel, N., Evaluation of survival data and two new rank order statistics arising in its consideration, Cancer Chemotherapy Rep. 50:163-170 (1966).

[30] Mantel, N. and Haenszel, N., Statistical aspects of the analysis of data for retrospective studies of disease, J. National Cancer Instit. 22:719-748 (1959).

[31] Marsh, G.M., OCMAP: a user oriented occupational cohort mortality analysis program, Amer. Statistician 34:245 (1980).

[32] McMichael, A.J., Standardized mortality ratios and the healthy worker effect: scratching beneath the surface, J. Occupational Med. 18:165-168 (1976).

[33] McMichael, A.J., Haynes, S.G., and Tyroler, H.A., Observations on the evaluation of occupational mortality data, J. Occupational Med. 17:128-131 (1975).

[34] Monson, R.R., Analysis of relative survival and proportional mortality, Computers and Biomedical Res. 1:325-332 (1974).

[35] Ortmeyer, C.E., Costello, J., Morgan, W.K.C., Swecker, S., and Peterson, M., The mortality of Appalachian coal miners, 1963 to 1971, Arch. of Environ. Health 29:67-72 (1974).

[36] Pell, S., O'Berg, M.T. and Karrh, B.W., Cancer epidemiology surveillance in the Du Pont Company, J. Occupational Med. 20:725-740 (1978).

[37] Scott, L.M., Bahler, K.W., DeLaGarza, A., and Lincoln, T.A., Mortality of uranium and non-uranium workers, Health Physics 23:555-557 (1972).

[38] Shindell, S., Weisberg, F.J., and Geifer, E.E., The "healthy worker effect"- Fact or artifact? J. Occupational Med. 20:807-811 (1978).

[39] Silcock, H., The comparison of occupational mortality rates, Population Studies 13:183-192 (1959).

[40] Symons, M.J. and Taulbee, J.D., Practical considerations for approximating relative risk by the standardized mortality ratio, J. Occupational Med. 23: 413-416 (1981).

[41] Symons, M.J. and Taulbee, J.D., Chapter 2: Statistical evaluation of the risk of cancer mortality among industrial populations. Statistical Methods for Cancer Studies, Cornell, R.G. (ed.). Marcel Dekker, Inc., New York (1984).

[42] Whittemore, A., Epidemiological implications of the multistage theory of carcinogenesis. Environmental Health: Quantitative Methods. Proceedings of a Conference on Environmental Health, Whittemore, A. (ed.). Philadelphia, SIAM: 72-87 (1977).

[43] Wong, O., Further criticisms on epidemiological methodology in occupational studies, J. Occupational Med. 19:220-222 (1977).

[44] Wong, O., Utidjian, M.D., and Karten, V.S., Retrospective evaluation of reproductive performance of workers exposed to ethylene dibromide, J. Occupational Med. 21:98-102 (1979).

[45] Woolson, R.F., Rank tests and a one-sample log rank test for comparing observed survival data to a standard population, Biometrics 37:687-698 (1981).

[46] Yule, G.U., On some points relating to vital statistics, more especially statistics of occupational mortality (with discussion), J. Roy. Statist. Soc. 94:1-84 (1934).

APPENDIX: Alternative Calculations of Expected Number of Deaths

Breslow (1978) shows that the probability of failure under study is equal to the expected value of the cumulative hazard. Specifically,

$$E(\delta_i) = E\{\int_0^{t_i} \lambda_i(u)du\}. \qquad (A1)$$

For standardized mortality calculations, each side of the equation (A1) has been utilized for computing the expected number of deaths.

In applications, inequality rather than equality exists between the probability of failure and the expected cumulative hazard. The critical step in the derivation of the equality of (A1) requires an interchange in the order of integration. Although not required to be known, the hazard of being lost to follow-up must be continuous for the order of integration to be interchanged. But for any finite length study the hazard of being censored has a discontinuity at the end of study, when all those alive are censored indicating an instantaneous, infinite increase in risk.

The two approaches used in calculating the number of expected deaths in a cohort mortality study are now described. Both methods require age-specific death rates for the cause of death of interest. The Modified Life Table Methods has been used almost exclusively in occupational cohort mortality studies. The description by Hill (1972) and the computer program implementations by Monson (1974) and Marsh (1980) are commonly used for applications to cohort mortality studies. Mechanically, the person-years of experience accumulated by the cohort of workers over the observation period in each age band multiplied by the corresponding age-specific death rate determines the number of expected deaths for the age band. Summing over all the age bands gives the total number of expected deaths.

The other approach, the Life Table Method, requires the use of life table conditional probabilities, but it has been less frequently applied to occupational mortality study data. For two such applications, see Scott, et al. (1972) and Ortmeyer, et al. (1974). Also in fertility studies the approach has recently been used by Wong, et al. (1979) and modified by Levine, et al. (1980). Essentially, the age-specific death rates are transformed into conditional probabilities of dying for the corresponding age band by one of the standard methods from current life table construction. Chiang (1972) presents the various conversions of rates to conditional probabilities of death that are utilized in current life table construction. The probability of dying during the observation period of the study can then be calculated for each individual. The sum of these probabilities over all individuals is taken as the total number of expected deaths.

The number of expected deaths calculated from the Modified Life Table Method is always larger than that calculated by the Life Table Method. The magnitude of the discrepancy in the two methods of calculating the number of expected deaths depends upon the cause of death being studied, the age distribution of the cohort, and the length of the study period. For a rare cause of death, such as a site specific cancer, the results from the two approaches are the same for all practical purposes. However, the calculation of number of expected deaths for all causes of death can depend dramatically on the method utilized. The older the age distribution of the cohort members and the longer the period of observation, the greater will be the discrepancy in the number of expected deaths computed by the two methods.

Algebraic demonstration of these results is straightforward. By examining series expansions of current life table conversions of specific rates into conditional probabilities of death, the numerical value for the probability of dying is readily seen to be exceeded by that of the corresponding rate. For example, the linear conversion of the age specific rate, R_x, to the conditional probability of death, q_x, during the age interval, indexed by x, is

$$q_x = R_x[1 + \tfrac{1}{2}R_x]^{-1} \qquad (A2)$$

Using the binomial to expand $(1 + \tfrac{1}{2}R_x)^{-1}$ in a series yields

$$q_x = R_x[1 - \tfrac{1}{2}R_x + (\tfrac{1}{2}R_x)^2 - (\tfrac{1}{2}R_x)^3 + \ldots]$$

and then

$$q_x = R_x - \tfrac{1}{2}R_x^2 + 1/4R_x^3 - 1/8R_x^4 + \ldots, \qquad (A3)$$

the series converging for R_x less than one. Since the age specific rates are positive,

$$q_x < R_x, \qquad (A4)$$

strictly, since $\tfrac{1}{2}R_x^2$ is positive. Consequently, for any age band the expected value of the right side of (A1) exceeds that of the left side of (A1). Then the division by these expected numbers in an SMR ratio of observed to expected and will produce an inequality in the opposite direction. The inequality also holds for two or more age bands, and hence the SMRs by the Life Table Method will exceed those by the Modified Life Table Method.

However, when the rate R_x is small so that R_x^2 and higher power terms are negligible, then

$$q_x \doteq R_x. \qquad (A5)$$

Notice that this is a numerical approximation; the probability of dying and the corresponding rate are two distinct concepts.

Several cohorts of employed populations were used by Bittinger (1982) that illustrate the practical differences in these two alternatives for computing the expected number of deaths. Table A1 contains the results for one example that shows a 20% relative discrepancy in the standardized mortality ratio for all causes of death computed using the expected numbers of death by the two methods. The magnitude of the differences possible depends upon the length of follow-up, the age distribution of the workforce, the size of the workforce and the cause of death under study. For more common causes of death, the death rates are larger, which is the single most important factor affecting the size of the discrepancy. Longer follow-up, older workers, and larger workforces also contribute to greater differences in these two computational strategies.

Table A1. Standardized Mortality Ratios by Two Methods of Calculating the Expected Numbers for Large Cohort

Cause of Death (number of deaths)	Modified Life Table Method SMR	Life Table Method SMR'	Ratio SMR/SMR'
Leukemia (25)	1.382	1.385	0.998
All Neoplasms (463)	0.976	1.018	0.959
Ischemic Heart Disease (1017)	0.954	1.056	0.904
All Causes (2373)	0.941	1.180	0.800

Adapted from Table 5 of Bittinger (1982).

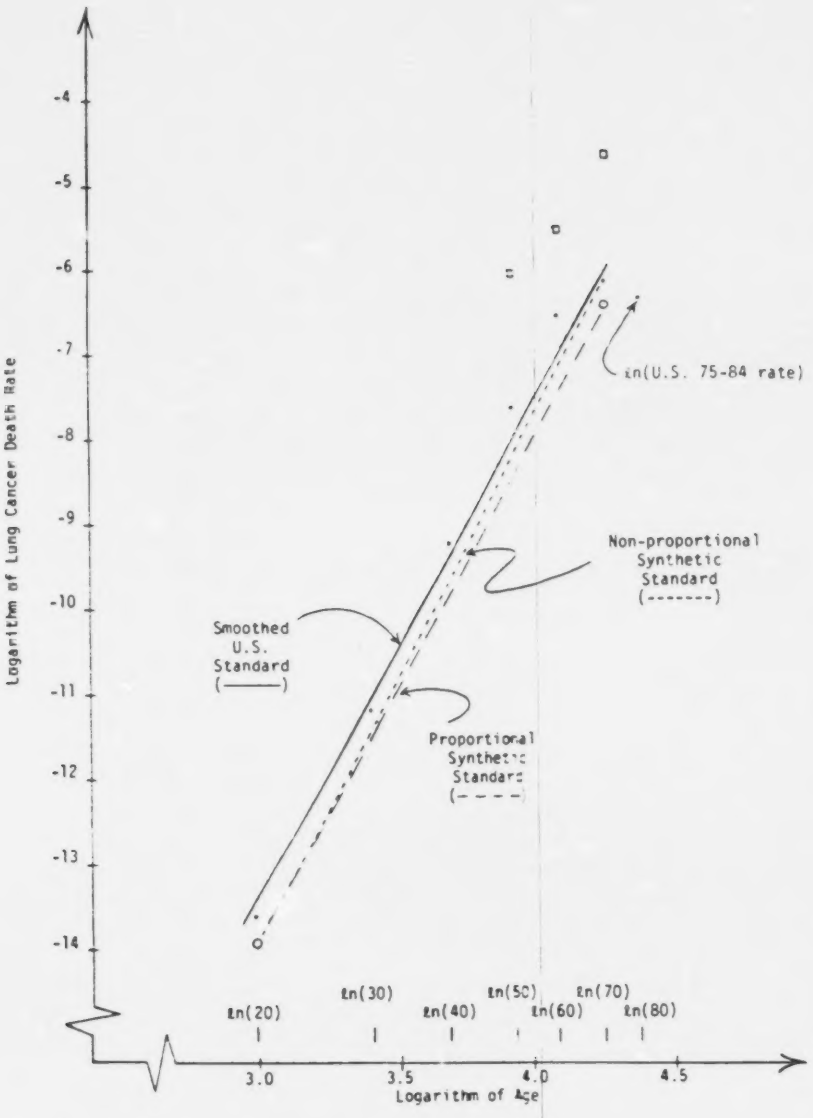


Figure 1. Smoothed U.S. Lung Cancer Rates, Synthetic Standards, and Observed Rates (□)